

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
OAKLAND DIVISION**

J. DOE 1, et al.,

Plaintiff(s),

vs.

GITHUB, INC., et al.,

Defendant(s).

Case Numbers: 4:22-cv-06823-JST  
4:22-cv-07074-JST

[PROPOSED] ORDER RE: DISCOVERY  
OF ELECTRONICALLY STORED  
INFORMATION

Upon the stipulation of the parties, the Court ORDERS as follows:

1. This Order supplements all other discovery rules and orders. It streamlines Electronically Stored Information (“ESI”) production to promote a “just, speedy, and inexpensive determination of this action, as required by Federal Rule of Civil Procedure 1.”

2. This Order may be modified in the Court’s discretion or by stipulation.

3. As in all cases, costs may be shifted for disproportionate ESI production requests pursuant to Federal Rule of Civil Procedure 26. Likewise, a party’s nonresponsive or dilatory discovery tactics are cost-shifting considerations.

4. A party’s meaningful compliance with this Order and efforts to promote efficiency and reduce costs will be considered in cost-shifting determinations.

5. The parties are expected to comply with the District’s E-Discovery Guidelines (“Guidelines”) and are encouraged to employ the District’s Model Stipulated Order Re: the Discovery of Electronically Stored Information and Checklist for Rule 26(f) Meet and Confer regarding Electronically Stored Information.

1           6. General ESI production requests under Federal Rules of Civil Procedure 34 and 45  
2 shall not include email or other forms of electronic correspondence or messaging service  
3 (collectively “email”). To obtain email, parties must propound specific email production  
4 requests.

5           7. Email production requests shall only be propounded for specific issues, rather than  
6 general discovery of a product or business.

7           8. Email production requests shall identify the custodian, search terms, and time frame.  
8 The parties shall cooperate to identify the proper custodians, proper search terms and proper  
9 timeframe as set forth in the Guidelines.

10          9. The parties will disclose at the outset any use of enterprise messaging tools, such as  
11 Slack or MS Teams, including the type of subscription and retention settings in place. The  
12 parties will meet and confer regarding issues relating to the parameters for collection of such  
13 data, the format of the production, and other unique issues to this type of data.

14          10. Each requesting party shall limit its email production requests to a total of five  
15 custodians per producing party for all such requests. OpenAI, Inc., OpenAI, L.P., OpenAI  
16 OpCo, L.L.C., OpenAI GP, L.L.C., OpenAI Startup Fund GP I, L.L.C., OpenAI Startup Fund I,  
17 L.P., and OpenAI Startup Fund Management, LLC shall be considered a single “producing  
18 party.” The parties may jointly agree to modify this limit without the Court’s leave. The Court  
19 shall consider contested requests for additional custodians, upon showing a distinct need based  
20 on the size, complexity, and issues of this specific case. Cost-shifting may be considered as part  
21 of any such request.

22          11. Each requesting party shall limit its email production requests to a total of five search  
23 terms per custodian per producing party. All search terms per custodian per producing party shall  
24 be identified in email production requests upon service. The parties may jointly agree to modify  
25 this limit without the Court’s leave. The Court shall consider contested requests for additional  
26 search terms per custodian, upon showing a distinct need based on the size, complexity, and  
27 issues of this specific case. The Court encourages the parties to confer on a process to test the  
28 efficacy of the search terms. The search terms shall be narrowly tailored to particular issues.

Indiscriminate terms, such as the producing company's name or its product name, are inappropriate unless combined with narrowing search criteria that sufficiently reduce the risk of overproduction. A conjunctive combination of multiple words or phrases (*e.g.*, "computer" and "system") narrows the search and shall count as a single search term. A disjunctive combination of multiple words or phrases (*e.g.*, "computer" or "system") broadens the search, and thus each word or phrase shall count as a separate search term unless they are variants of the same word. Use of narrowing search criteria (*e.g.*, "and," "but not," "w/x") is encouraged to limit the production and shall be considered when determining whether to shift costs for disproportionate discovery. Should a party serve email production requests with search terms beyond the limits agreed to by the parties or granted by the Court pursuant to this paragraph, this shall be considered in determining whether any party shall bear all reasonable costs caused by such additional discovery.

12. Nothing in this Order prevents the parties from agreeing to use technology assisted review and other techniques insofar as their use improves the efficacy of discovery. Such topics should be discussed pursuant to the District's E-Discovery Guidelines.

### 13. Production Format for ESI.

a. Except as otherwise provided in this Order, all documents existing in electronic format (*e.g.*, Word, PDF, webpage files) shall be produced as 300 DPI TIFF black and white images with a Bates number stamped on each page, document metadata (including extracted or OCR text), and a link to the associated extracted or OCR text file. The parties will consider re-producing select and specifically identified Word documents as a color PDF upon reasonable request by the receiving party.

b. Source code and certain other highly confidential technical materials will be produced in native format and shall be made available for inspection on a secure computer in a secured room without Internet access or network access to other computers at a secure location determined by the producing party, pursuant to the terms of the Amended Protective Order (ECF

No. 145) governing “HIGHLY CONFIDENTIAL – SOURCE CODE” materials.

- c. Spreadsheet-type files (e.g., Microsoft Excel) and PowerPoint presentations should be produced in native format unless they require redaction.

Documents produced in native format should be named according to the Bates number assigned, with the Confidentiality designation appended to the Bates numbers with a hyphen. Single-page Bates-stamped TIFF image slip-sheets will be included for each document produced in native format. The slip-sheets will display the Bates number of the native file, the Confidentiality endorsement, and an endorsement stating, “File Produced Natively.”

- d. All attachments, addendums, enclosures, and/or exhibits to a parent document will be, to the extent reasonably possible, produced immediately following the parent document and identified as they relate to the respective parent document with the parent/child relationship intact.

- e. To the extent reasonably possible, ESI items shall be processed so as to preserve the date/time shown in the document as it was last saved, not the date of collection or processing.

- f. Telemetry data, training data, and other data inaccessible within the meaning of Federal Rule of Civil Procedure 26, shall not be produced to another party, either in native form or otherwise. The parties shall meet and confer in good faith regarding reasonable methodologies to query or sample such data sources. Given the large volume and nature of this data, searching the entirety or even portions of such data is impossible, complex, limited, and/or costly.

14. Metadata. Load or DAT files should include, where applicable and to the extent it exists and was captured at the time of the collection, the information listed in **Exhibit A** attached. To the extent that metadata does not exist or is not reasonably accessible or available

1 for any documents produced, nothing in this Order shall require any party to extract, capture,  
2 collect or produce such metadata.

3 15. Privilege Logs. A party that withholds documents, ESI, and any other materials and  
4 information on the grounds of attorney-client privilege, attorney work product, and/or any other  
5 applicable privilege, immunity or protection from discovery, shall produce a privilege log.  
6 Privileged communications exclusively between a party and its outside or in-house counsel on or  
7 after November 3, 2022, and work product performed by a party at the direction of its outside or  
8 in-house counsel on or after November 3, 2022 need not be included on a privilege log.  
9 Privilege logs shall be served no later than 60 days before the close of fact discovery. The  
10 privilege log shall include:

- 11 a. A unique number for each entry on the log.
- 12 b. Date of document for all ESI and to the extent known or discernible for all  
13 hard copy documents. For emails this should be the sent date of the  
14 document and for loose ESI this should be the last-modified or create date  
15 of the document.
- 16 c. Custodian of the document. For emails this should be populated with the  
17 metadata extracted from the Custodian(s) field. For loose ESI, this should  
18 be populated with the Custodian value assigned to the instance of the  
19 document in question.
- 20 d. Author of the document. For emails this should be populated with the  
21 metadata extracted from the "Email From" field associated with the file.  
22 For loose ESI, this should be populated with the metadata extracted from the  
23 "Author" field. For hard copy documents, this will be populated with the  
24 "Custodian."
- 25 e. Recipient(s) of the document where reasonably ascertainable. For emails  
26 this will be populated with the metadata extracted from the "Email To" field  
27 associated with the file. Separate columns will be included for the metadata  
28 extracted from the "Email CC" and "Email BCC" fields, where populated.

1 f. The type of privilege being asserted, for example: (a) AC for  
2 Attorney/Client, (b) WP for Attorney Work Product, (c) CI for Common  
3 Interest.

4 g. Nothing in Paragraph 15 of this Order shall be read to preclude the parties  
5 from logging documents categorically where appropriate.

6 16. Any Party that issues a non-party subpoena shall timely notify other Parties when it  
7 receives non-party productions of documents, ESI or other responses from the non-party, and  
8 shall provide copies of all such productions in the format in which they were received from the  
9 non-party to the other Parties. If the non-party produces documents or ESI that are not Bates  
10 labeled, the Receiving Party will Bates-label the documents or ESI before producing a copy to  
11 the other Parties if the produced documents or ESI are of the type so requiring consistent with  
12 other provisions of this Order. Nothing in this Protocol precludes a Receiving Party from  
13 reviewing a non-party production of documents or ESI and applying (if the non-party produced  
14 documents without a confidentiality designation) or elevating a confidentiality designation (e.g.,  
15 elevating documents designated as CONFIDENTIAL to ATTORNEYS' EYES ONLY) if the  
16 production contains confidential or extremely sensitive information as defined by the Amended  
17 Protective Order (ECF No. 145).

18  
19  
20 **IT IS ORDERED** that the forgoing is approved.

21 Dated:

22 \_\_\_\_\_  
23 Honorable Jon S. Tigar  
24 United States District Judge  
25  
26  
27  
28

**EXHIBIT A**

<i>Field</i>	<i>Data Type</i>	<i>Paper</i>	<i>Native Files &amp; Email Attachments</i>	<i>Email</i>
BegDoc	Integer - Text	Starting Bates #	Starting Bates #	Starting Bates #
EndDoc	Integer - Text	Ending Bates #	Ending Bates #	Ending Bates #
BegAttach	Integer - Text	Starting bates # of document family	Starting bates # of document family	Starting bates # of document family
EndAttach	Integer - Text	Ending bates # of document family	Ending bates # of document family	Ending bates # of document family
Custodians	Text – paragraph Separate entries with “,”		All names of people the document was collected from even if removed from production as a duplicate	All names of people the document was collected from even if removed from production as a duplicate
From	Text - paragraph			Sender of message
To	Text – paragraph Separate entries with “,”			Recipients of message
CC	Text – paragraph Separate entries with “,”			Copied recipients
BCC	Text – paragraph Separate entries with “,”			Blind copied recipients
Subject	Text - paragraph			Subject of message
DateSent	Date (mm/dd/yy yy)			Date message sent
DateReceived	Date (mm/dd/yy yy)			Date message received
FileName	Text -		Name of original	Name of original

<i>Field</i>	<i>Data Type</i>	<i>Paper</i>	<i>Native Files &amp; Email Attachments</i>	<i>Email</i>
	paragraph		file including extension	file including extension
FileExtension	Text		Extension of original file	Extension of original file
DateCreated	Date/Time (mm/dd/yy yy)		Date file was created	
DateModified	Date/Time (mm/dd/yy yy)		Date file was last modified	
Author	Text - paragraph		Document author from metadata	
Confidentiality	Single choice	Designation branded onto the document	Designation branded onto the document	Designation branded onto the document
FullText	Link	Link to the extracted or OCR text file	Link to the extracted or OCR text file	Link to the extracted or OCR text file
DocLink	Link	Link to the native file	Link to the native file	Link to the native file
MD5Hash	Text		The MD5Hash value of the file	The MD5Hash value of the file